

CLAIMS

What is claimed is:

1. A method for evaluating similarity among a plurality of data structures, comprising:
2. analyzing each structure of said plurality of data structures to generate
3. at least one substructure;
4. matching said at least one substructure to a database having a plurality |
5. of entries to obtain at least one matching entry; and
6. generating a match value using a relative entropy value corresponding
7. to said at least one matching entry.

1. 2. The method according to claim 1, further comprising:
2. creating said plurality of entries in said database; and
3. processing said plurality of entries in said database.

1. 3. The method according to claim 2, wherein said creating further
2. comprises creating said plurality of entries using a tool having a graphical user
3. interface and exporting said plurality of entries to said database.

1. 4. The method according to claim 2, wherein said processing further
2. comprises:
3. verifying said plurality of entries for validity; and
4. calculating said relative entropy value corresponding to each entry of
5. said plurality of entries.

1 5. The method according to claim 4, wherein said processing further
2 comprises storing said each entry of said plurality of entries together with said
3 corresponding relative entropy value in a compressed format.

1 6. The method according to claim 1, further comprising extracting
2 from a lexicon database having a plurality of elements each element associated
3 to said each structure, assigning at least one code of said each element to said
4 each structure, and retrieving said at least one code during matching to obtain
5 said at least one matching entry.

1 7. The method according to claim 6, further comprising reading
2 lexical probability files and assigning a probability value to said each element of
3 said plurality of elements in said lexicon database.

1 8. The method according to claim 1, wherein each structure of said
2 plurality of data structures is a representation of a linguistic expression.

1 9. The method according to claim 4, wherein said database is a
2 thesaurus hierarchy including a root entry, said plurality of entries depending
3 from said root entry.

1 10. The method according to claim 9, wherein said relative entropy
2 value corresponding to said each entry of said plurality of entries is calculated
3 based on an entropy value of said each entry and an entropy value of said root
4 entry.

1 11. The method according to claim 6, wherein said each element in
2 said lexicon database is a word.

1 12. A method for evaluating similarity among a plurality of data
2 structures comprising:
3 creating a plurality of entries in a database; and
4 calculating a relative entropy value corresponding to each entry of said
5 plurality of entries.

1 13. The method according to claim 12, further comprising storing said
2 each entry of said plurality of entries together with said corresponding relative
3 entropy value in a compressed format.

1 14. The method according to claim 12, further comprising:
2 creating said plurality of entries using a tool having a graphical user
3 interface; and
4 exporting said plurality of entries to said database.

1 15. The method according to claim 12 further comprising:
2 analyzing each structure of said plurality of data structures to generate
3 at least one substructure;
4 matching said at least one substructure of said each structure to said
5 database to obtain at least one matching entry; and
6 generating a match value using said relative entropy value
7 corresponding to said at least one matching entry.

1 16. The method according to claim 15, further comprising:

2 verifying said plurality of entries for validity;
3 extracting from a lexicon database having a plurality of elements each
4 element associated to said each structure;
5 reading lexical probability files;
6 assigning a probability value to said each element of said plurality of
7 elements in said lexicon database;
8 assigning at least one code of said each element to said each structure;
9 and
10 retrieving and matching said at least one code to said database to obtain
11 said at least one matching entry.

1 17. The method according to claim 16, wherein said each structure of
2 said plurality of data structures is a representation of a linguistic expression.

1 18. The method according to claim 12, wherein said database is a
2 thesaurus hierarchy including a root entry, said plurality of entries depending
3 from said root entry.

1 19. The method according to claim 18, wherein said relative entropy
2 value corresponding to said each entry of said plurality of entries is calculated
3 based on an entropy value for said each entry and an entropy value for said
4 root entry.

1 20. The method according to claim 15, wherein said each element in
2 said lexicon database is a word.

1 21. A computer readable medium containing executable instructions
2 which, when executed in a processing system, cause the system to perform a
3 method for evaluating similarity among a plurality of data structures, the
4 method comprising:
5 analyzing each structure of said plurality of data structures to generate
6 at least one substructure;
7 matching said at least one substructure to a database having a plurality
8 of entries to obtain at least one matching entry; and
9 generating a match value using a relative entropy value corresponding
10 to said at least one matching entry.

1 22. The computer readable medium according to claim 21, wherein
2 the method further comprises:
3 creating said plurality of entries in said database; and
4 processing said plurality of entries in said database.

1 23. The computer readable medium according to claim 22, wherein
2 said creating further comprises creating said plurality of entries using a tool
3 having a graphical user interface and exporting said plurality of entries to said
4 database.

1 24. The computer readable medium according to claim 22, wherein
2 said processing further comprises:
3 verifying said plurality of entries for validity; and
4 calculating said relative entropy value corresponding to each entry of
5 said plurality of entries.

1 25. The computer readable medium according to claim 24, wherein
2 said processing further comprises storing said each entry of said plurality of
3 entries together with said corresponding relative entropy value in a
4 compressed format.

1 26. The computer readable medium according to claim 21, further
2 comprising extracting from a lexicon database having a plurality of elements
3 each element associated to said each structure, assigning at least one code of
4 said each element to said each structure, and retrieving said at least one code
5 during matching to obtain said at least one matching entry.

1 27. The computer readable medium according to claim 26, further
2 comprising reading lexical probability files and assigning a probability value to
3 said each element of said plurality of elements in said lexicon database.

1 28. The computer readable medium according to claim 21, wherein
2 each structure of said plurality of data structures is a representation of a
3 linguistic expression.

1 29. The computer readable medium according to claim 24, wherein
2 said database is a thesaurus hierarchy including a root entry, said plurality of
3 entries depending from said root entry.

1 30. The computer readable medium according to claim 29, wherein
2 said relative entropy value corresponding to said each entry of said plurality of
3 entries is calculated based on an entropy value of said each entry and an
4 entropy value of said root entry.

1 31. The computer readable medium according to claim 26, wherein
2 said each element in said lexicon database is a word.

1 32. A computer readable medium containing executable instructions
2 which, when executed in a processing system, cause the system to perform a
3 method for evaluating similarity among a plurality of data structures, the
4 method comprising:
5 creating a plurality of entries in a database; and
6 calculating a relative entropy value corresponding to each entry of said
7 plurality of entries.

1 33. The computer readable medium according to claim 32, further
2 comprising storing said each entry of said plurality of entries together with said
3 corresponding relative entropy value in a compressed format.

1 34. The computer readable medium according to claim 32, further
2 comprising:
3 creating said plurality of entries using a tool having a graphical user
4 interface; and
5 exporting said plurality of entries to said database.

1 35. The computer readable medium according to claim 32 further
2 comprising:
3 analyzing each structure of said plurality of data structures to generate
4 at least one substructure;
5 matching said at least one substructure of said each structure to said
6 database to obtain at least one matching entry; and

7 generating a match value using said relative entropy value
8 corresponding to said at least one matching entry.

1 36. The computer readable medium according to claim 35, further
2 comprising:
3 verifying said plurality of entries for validity;
4 extracting from a lexicon database having a plurality of elements each
5 element associated to said each structure;
6 reading lexical probability files;
7 assigning a probability value to said each element of said plurality of
8 elements in said lexicon database;
9 assigning at least one code of said each element to said each structure;
10 and
11 retrieving and matching said at least one code to said database to obtain
12 said at least one matching entry.

1 37. The computer readable medium according to claim 36, wherein
2 said each structure of said plurality of data structures is a representation of a
3 linguistic expression.

1 38. The computer readable medium according to claim 32, wherein
2 said database is a thesaurus hierarchy including a root entry, said plurality of
3 entries depending from said root entry.

1 39. The computer readable medium according to claim 38, wherein
2 said relative entropy value corresponding to each entry of said plurality of

3 entries is calculated based on an entropy value for said each entry and an
4 entropy value for said root entry.

1 40. The computer readable medium according to claim 35, wherein
2 said each element in said lexicon database is a word.

1 41. An article of manufacture comprising a program storage medium
2 readable by a computer and tangibly embodying at least one program of
3 instructions executable by said computer to perform method steps for
4 evaluating similarity among a plurality of data structures, said method
5 comprising:

6 analyzing each structure of said plurality of data structures to generate
7 at least one substructure;
8 matching said at least one substructure to a database having a plurality
9 of entries to obtain at least one matching entry; and
10 generating a match value using a relative entropy value corresponding
11 to said at least one matching entry.

1 42. The article of manufacture according to claim 41, wherein the
2 method further comprises:

3 creating said plurality of entries in said database; and
4 processing said plurality of entries in said database.

1 43. The article of manufacture according to claim 42, wherein said
2 creating further comprises creating said plurality of entries using a tool having
3 a graphical user interface and exporting said plurality of entries to said
4 database.

1 44. The article of manufacture according to claim 42, wherein said
2 processing further comprises:
3 verifying said plurality of entries for validity; and
4 calculating said relative entropy value corresponding to each entry of
5 said plurality of entries.

1 45. The article of manufacture according to claim 44, wherein said
2 processing further comprises storing each entry of said plurality of entries
3 together with said corresponding relative entropy value in a compressed
4 format.

1 46. The article of manufacture according to claim 41, wherein the
2 method further comprises:
3 extracting from a lexicon database having a plurality of elements each
4 element associated to said each structure;
5 assigning at least one code of said each element to said each structure;
6 and
7 retrieving said at least one during matching to obtain said at least one
8 matching entry.

1 47. The article of manufacture according to claim 46, wherein the
2 method further comprises reading lexical probability files and assigning a
3 probability value to said each element of said plurality of elements in said
4 lexicon database.

1 48. The article of manufacture according to claim 41, wherein each
2 structure of said plurality of data structures is a representation of a linguistic
3 expression.

1 49. The article of manufacture according to claim 44, wherein said
2 database is a thesaurus hierarchy including a root entry, said plurality of entries
3 depending from said root entry.

1 50. The article of manufacture according to claim 49, wherein said
2 relative entropy value corresponding to said each entry of said plurality of
3 entries is calculated based on an entropy value of said each entry and an
4 entropy value of said root entry.

1 51. The article of manufacture according to claim 46, wherein said
2 each element in said lexicon database is a word.

1 52. An article of manufacture comprising a program storage medium
2 readable by a computer and tangibly embodying at least one program of
3 instructions executable by said computer to perform method steps for
4 evaluating similarity among a plurality of data structures, said method

5 comprising:

6 creating a plurality of entries in a database; and

7 calculating a relative entropy value corresponding to each entry of said
8 plurality of entries.„

1 53. The article of manufacture according to claim 52, wherein the
2 method further comprises storing said each entry of said plurality of entries

3 together with said corresponding relative entropy value in a compressed
4 format.

1 54. The article of manufacture according to claim 52, wherein the
2 method further comprises:

3 creating said plurality of entries using a tool having a graphical user
4 interface; and
5 exporting said plurality of entries to said database.

1 55. The article of manufacture according to claim 52, wherein the
2 method further comprises:

3 analyzing each structure of said plurality of data structures to generate
4 at least one substructure;
5 matching said at least one substructure of said each structure to said
6 database to obtain at least one matching entry; and
7 generating a match value using said relative entropy value
8 corresponding to said at least one matching entry.

1 56. The article of manufacture according to claim 55, wherein the
2 method further comprises:

3 verifying said plurality of entries for validity;
4 extracting from a lexicon database having a plurality of elements each
5 element associated to said each structure;
6 reading lexical probability files;
7 assigning a probability value to said each element of said plurality of
8 elements in said lexicon database;

9 assigning at least one code of said each element to said each structure;
10 and
11 retrieving and matching said at least one code to said database to obtain
12 said at least one matching entry.

1 57. The article of manufacture according to claim 56, wherein said
2 structure of said plurality of data structures is a representation of an linguistic
3 expression.

1 58. The article of manufacture according to claim 52, wherein said
2 database is a thesaurus hierarchy including a root entry, said plurality of entries
3 depending from said root entry.

1 59. The article of manufacture according to a claim 58, wherein said
2 relative entropy value corresponding to said each entry of said plurality of
3 entries is calculated based on an entropy value for said each entry and an
4 entropy value for said root entry.

1 60. The article of manufacture according to claim 55, wherein said
2 each element in said lexicon database is a word.

1 61. A system for evaluating similarity among a plurality of data
2 structures, comprising:
3 means for analyzing each structure of said plurality of data structures to
4 generate at least one substructure;
5 means for matching said at least one substructure to a database having a
6 plurality of entries to obtain at least one matching entry; and

7 means for generating a match value using a relative entropy value
8 corresponding to said at least one matching entry.

1 62. The system according to claim 61, further comprising:
2 means for creating said plurality of entries in said database; and
3 means for processing said plurality of entries in said database.

1 63. The system according to claim 62, wherein said creating means
2 further comprises means for creating said plurality of entries using a tool
3 having a graphical user interface and exporting said plurality of entries to said
4 database.

1 64. The system according to claim 62, wherein said processing means
2 further comprises:
3 means for verifying said plurality of entries for validity; and
4 means for calculating said relative entropy value corresponding to each
5 entry of said plurality of entries.

1 65. The system according to claim 64, wherein said processing means
2 further comprises means for storing said each entry of said plurality of entries
3 together with said corresponding relative entropy value in a compressed
4 format.

1 66. The system according to claim 61, further comprising:
2 means for extracting from a lexicon database having a plurality of
3 elements each element associated to said each structure;

4 means for assigning at least one code of said each element to said each
5 structure; and
6 means for retrieving said at least one code during matching to obtain
7 said at least one matching entry.

1 67. The system according to claim 66, further comprising:
2 means for reading lexical probability files; and
3 means for assigning a probability value to said each element of said
4 plurality of elements in said lexicon database.

1 68. The system according to claim 61, wherein each structure of said
2 plurality of data structures is a representation of a linguistic expression.

1 69. The system according to claim 64, wherein said database is a
2 thesaurus hierarchy including a root entry, said plurality of entries depending
3 from said root entry.

1 70. The system according to claim 69, wherein said relative entropy
2 value corresponding to said each entry of said plurality of entries is calculated
3 based on an entropy value of said each entry and an entropy value of said root
4 entry.

1 71. The system according to claim 66, wherein said each element in
2 said lexicon database is a word.

1 72. A system for evaluating similarity among a plurality of data
2 structures, comprising:

3 means for creating a plurality of entries in a database; and
4 means for calculating a relative entropy value corresponding to each
5 entry of said plurality of entries.✓

1 73. The system according to claim 72, further comprising means for
2 storing said each entry of said plurality of entries together with said
3 corresponding relative entropy value in a compressed format.

1 74. The system according to claim 72, further comprising:
2 means for creating said plurality of entries using a tool having a
3 graphical user interface; and
4 means for exporting said plurality of entries to said database.

1 75. The system according to claim 72, further comprising:
2 means for analyzing each structure of said plurality of data structures to
3 generate at least one substructure;
4 means for matching said at least one substructure of said each structure
5 to said database to obtain at least one matching entry; and
6 means for generating a match value using said relative entropy value
7 corresponding to said at least one matching entry.

1 76. The system according to claim 75, further comprising:
2 means for verifying said plurality of entries for validity;
3 means for extracting from a lexicon database having a plurality of
4 elements each element associated to said each structure;
5 means for reading lexical probability files;

6 means for assigning a probability value to said each element of said
7 plurality of elements in said lexicon database;
8 means for assigning at least one code of said each element to said each
9 structure; and
10 means for retrieving and matching said at least one code to said database
11 to obtain said at least one matching entry.

1 77. The system according to claim 76, wherein said each structure of
2 said plurality of data structures is a representation of a linguistic expression.

1 78. The system according to claim 72, wherein said database is a
2 thesaurus hierarchy including a root entry, said plurality of entries depending
3 from said root entry.

1 79. The system according to claim 78, wherein said relative entropy
2 value corresponding to said each entry of said plurality of entries is calculated
3 based on an entropy value for said each entry and an entropy value for said
4 root entry.

1 80. The system according to claim 75, wherein said each element in
2 said lexicon database is a word.

1 81. A system for evaluating similarity among a plurality of data
2 structures, comprising:
3 a database having a plurality of entries;

4 an analyzer, coupled to said database, said analyzer configured to
5 analyze each structure of said plurality of data structure to generate at least one
6 substructure;
7 a matching unit, coupled to said analyzer and said database, said
8 matching unit configured to match said at least one substructure to at least one
9 entry of said plurality of entries to obtain at least one matching entry; and
10 an entropy calculator, coupled to said matching unit and said database,
11 configured to generate a match value using a relative entropy value
12 corresponding to said at least one matching entry.

1 82. The system according to claim 81, wherein said plurality of entries
2 are created offline using a tool having a graphical user interface and are
3 exported to said database.

1 83. The system according to claim 81, wherein said entropy calculator
2 further calculates said relative entropy value corresponding to each entry of
3 said plurality of entries.

1 84. The system according to claim 83, wherein said database stores
2 said each entry together with said corresponding relative entropy value in a
3 compressed format.

1 85. The system according to claim 81, wherein said matching unit
2 further retrieves at least one code from said at least one substructure and
3 matches said at least one code to said at least one entry to obtain said at least
4 one matching entry.

1 86. The system according to claim 81, wherein each structure of said
2 plurality of data structures is a representation of a linguistic expression.

1 87. The system according to claim 81, wherein said database is a
2 thesaurus hierarchy including a root entry, said plurality of entries depending
3 from said root entry.

1 88. The system according to claim 87, wherein said relative entropy
2 value corresponding to said each entry of said plurality of entries is calculated
3 based on an entropy value of said each entry and an entropy value of said root
4 entry.